

PCT / 1-B 04 / 01010



Europäisches
Patentamt

European
Patent Office

Office européen
des brevets

REC'D 13 APR 2004

WIPO

PCT

Bescheinigung

Certificate

Attestation

Die angehefteten Unterla-
gen stimmen mit der
ursprünglich eingereichten
Fassung der auf dem näch-
sten Blatt bezeichneten
europäischen Patentanmel-
dung überein.

The attached documents
are exact copies of the
European patent application
described on the following
page, as originally filed.

Les documents fixés à
cette attestation sont
conformes à la version
initialement déposée de
la demande de brevet
européen spécifiée à la
page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

03290852.7

**PRIORITY
DOCUMENT**
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk



Anmeldung Nr:
Application no.: 03290852.7
Demande no:

Anmeldetag:
Date of filing: 04.04.03
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Koninklijke Philips Electronics N.V.
Groenewoudseweg 1
5621 BA Eindhoven
PAYS-BAS

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se référer à la description.)

Video encoding method and corresponding encoding and decoding devices

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s)
revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

H04N7/20

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of
filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LU MC NL
PT RO SE SI SK TR LI

"VIDEO ENCODING METHOD AND CORRESPONDING ENCODING AND DECODING DEVICES"**FIELD OF THE INVENTION**

5 The present invention relates to the field of video compression and, for instance, to the video coding standards of the MPEG family (MPEG-1, MPEG-2, MPEG-4) and to the video recommendations of the ITU-H.26X family (H.261, H.263 and extensions, H.264). More specifically, this invention concerns an encoding method applied to an input video sequence corresponding to successive scenes subdivided into successive video object planes (VOPs) and generating, for coding
10 all the video objects of said scenes, a coded bitstream constituted of encoded video data in which each data item is described by means of a bitstream syntax allowing to recognize and decode all the elements of the content of said bitstream, said content being described in terms of separate channels.

15 The invention also relates to a corresponding encoding device, to a transmittable video signal consisting of a coded bitstream generated by such an encoding device, and to a method and a device for decoding a video signal consisting of such a coded bitstream.

BACKGROUND OF THE INVENTION

20 In the first video coding standards and recommendations (up to MPEG-4 and H.264), the video was assumed to be rectangular and to be described in terms of a luminance channel and two chrominance channels. With MPEG-4, an additional channel carrying shape information has been introduced. Two modes are available to compress those channels : the INTRA mode, where each channel is encoded by exploiting the spatial redundancy of the pixels in a given channel of a single image, and the INTER mode, exploiting the temporal redundancy between separate images. The INTER mode relies on a motion-compensation technique, which describes an image from one
25 (or more) previously decoded image(s) by encoding the motion of pixels from one image to the other. Usually, the image to be encoded is partitioned into independent blocks, each of them being assigned a motion vector. A prediction of the image is then constructed by displacing pixel blocks from the reference image(s) according to the set of motion vectors (luminance and chrominance channels share the same motion description). Finally, the difference (called the residual signal)
30 between the image to be encoded and its motion-compensated prediction is encoded in the INTER mode to further refine the decoded image.

35 The fact that all pixel channels are described by the same motion information is a limitation damaging the compression efficiency of the video coding system. In fact, several situations where the encoding of a motion vector set for all channels is not necessary can be identified. For instance, in sequences where there is little motion between successive frames, instead of encoding a full set of motion vectors repeating that each macroblock has no motion, it may be advantageous to signal that no motion is present. In other situations, instead of encoding a motion vector field, it may be advantageous to signal that the prediction of the motion vectors

should be constructed by interpolating the image from several reference images (in this case, the decoder has to estimate a motion field between several reference images and interpolate it to create the prediction of the current image), or a motion vector field can still be interpreted not directly from one or several reference image(s), but instead from the temporal interpolation of the reference images.

There are still some situations where the way of constructing the temporal prediction can be switched on a channel or macroblock basis. A first example is a sequence with a shape channel : it is possible that the shape information does not change much, whereas the luminance and chrominance channels carry varying information (it is for instance the case with a video depicting a rotating planet : the shape is always a disc, but the texture of it depends on the planet rotation). In this situation, the shape channel can be recovered by temporal copy, and the luminance and chrominance channels by motion compensated temporal interpolation. A second example is the case of a change at the macroblock level. In a video sequence showing a seascape with the sky in the upper part of the picture, unlike the sea, the sky remains the same from one image to the other. Its macroblocks can therefore be encoded by temporal copy, whereas the macroblocks of the sea have to be encoded by temporal interpolation.

SUMMARY OF THE INVENTION

It is therefore the object of the invention to propose a video encoding method in which it is provided to adapt the way the temporal prediction is formed.

To this end, the invention relates to a method such as defined in the introductory part of the description and which is moreover characterized in that said syntax comprises an additional syntactic element provided for describing independently the type of temporal prediction of the various channels, said predictions being chosen within a list comprising the following situations :

- the temporal prediction is formed by directly applying the motion field sent by the encoder on one or more reference pictures ;
- the temporal prediction is a copy of a reference image ;
- the temporal prediction is formed by the temporal interpolation of the motion field ;
- the temporal prediction is formed by the temporal interpolation of the current motion field and further refined by the motion field sent by the encoder.

The invention also relates to a corresponding encoding device, to a transmittable video signal consisting of a coded bitstream generated by such an encoding device, and to a method and a device for decoding a video signal consisting of such a coded bitstream.

DETAILED DESCRIPTION OF THE INVENTION

According to the invention, it is proposed to introduce in the encoding syntax used by the video standards and recommendations a new syntactic element supporting their lack of flexibility and opening new possibilities to encode more efficiently and independently the temporal

prediction of various channels. This additional syntactic element, called for instance "channel temporal prediction", takes the following symbolic values :

Motion_compensation

Temporal_copy

5 Temporal_interpolation

Motion_compensated_temporal_interpolation,

and the meaning of these values is :

10 a) motion_compensation : the temporal prediction is formed by directly applying the motion field sent by the encoder on one or more reference pictures (this default mode is implicitly the INTER coding mode of most of the current coding systems) ;

b) temporal_copy : the temporal prediction is a copy of a reference image ;

c) temporal_interpolation : the temporal prediction is formed by the temporal interpolation of the motion fields ;

15 d) motion_compensated_temporal_interpolation : the temporal prediction is formed by the temporal interpolation of the current motion field and further refined by the motion field sent by the encoder.

20 The words "temporal interpolation", must be understood in a broad sense, i.e. as meaning any operation of the type defined by an expression such as $V_{new} = a.V_{past} + b.V_{future} + K$, where V_{past} designates one (at least one) previous motion field, V_{future} designates one (at least one) future motion field, a and b designate coefficients respectively assigned to said past and future motion fields, K designates an offset and V_{new} is the new motion field thus obtained. It can therefore be seen that, in fact, the particular case of the temporal copy is included in the more general case of the temporal interpolation, for $b = 0$ and $K = 0$ (or $a = 0$ and $K = 0$).

25 The additional syntactic element thus proposed is expected to be placed at various levels in the coded bitstream that has to be stored, or transmitted to the decoding side ;

1) at the image level (or VOP level in MPEG-4 terminology), one syntactic element being placed in an INTER picture (its meaning is then shared by all the channels present in the VOP) ;

30 2) at the image level (or VOP level), with a syntactic element for each present channel ;

3) at the macroblock level, with moreover a specific syntactic element for each present channel.

35

CLAIMS :

1. An encoding method applied to an input video sequence corresponding to successive scenes subdivided into successive video object planes (VOPs) and generating, for coding all the video objects of said scenes, a coded bitstream constituted of encoded video data in which each data item is described by means of a bitstream syntax allowing to recognize and decode all the elements of the content of said bitstream, said content being described in terms of separate channels, said method being further characterized in that said syntax comprises an additional syntactic element provided for describing independently the type of temporal prediction of the various channels, said predictions being chosen within a list comprising the following situations :
- the temporal prediction is formed by directly applying the motion field sent by the encoder on one or more reference pictures ;
 - the temporal prediction is a copy of a reference image ;
 - the temporal prediction is formed by the temporal interpolation of the motion field ;
 - the temporal prediction is formed by the temporal interpolation of the current motion field and further refined by the motion field sent by the encoder.
2. An encoding method according to claim 1, characterized in that said additional syntactic element is placed at the image level in said generated coded bitstream and its meaning is shared by all existing channels.
3. An encoding device processing an input video sequence that corresponds to successive scenes subdivided into successive video object planes (VOPs) and generating, for coding all the video objects of said scenes, a coded bitstream constituted of encoded video data in which each data item is described by means of a bitstream syntax allowing to recognize and decode all the elements of the content of said bitstream, said content being described in terms of separate channels, said encoding device being provided for carrying out the encoding method according to anyone of claims 1 and 2.
4. A transmittable video signal consisting of a coded bitstream generated by an encoding device processing an input video sequence that corresponds to successive scenes subdivided into successive video object planes (VOPs) and generating, for coding all the video objects of said scenes, a coded bitstream constituted of encoded video data in which each data item is described by means of a bitstream syntax allowing to recognize and decode all the elements of the content of said bitstream, said content being described in terms of separate channels, said transmittable video signal including an additional syntactic element provided for describing independently the type of temporal prediction of the various channels, said predictions being chosen within a list comprising the following situations :
- the temporal prediction is formed by directly applying the motion field sent by the encoder on one or more reference pictures ;
 - the temporal prediction is a copy of a reference image ;
 - the temporal prediction is formed by the temporal interpolation of the motion field ;
 - the temporal prediction is formed by the temporal interpolation of the current

motion field and further refined by the motion field sent by the encoder.

5. A method for decoding a transmittable video signal consisting of a coded bitstream generated by an encoding device processing an input video sequence that corresponds to successive scenes subdivided into successive video object planes (VOPs) and generating, for coding all the video objects of said scenes, a coded bitstream constituted of encoded video data in which each data item is described by means of a bitstream syntax allowing to recognize and decode all the elements of the content of said bitstream, said content being described in terms of separate channels, said transmittable video signal including an additional syntactic element provided for describing independently the type of temporal prediction of the various channels, said predictions being chosen within a list comprising the following situations :

- the temporal prediction is formed by directly applying the motion field sent by the encoder on one or more reference pictures ;
- the temporal prediction is a copy of a reference image ;
- the temporal prediction is formed by the temporal interpolation of the motion field ;
- the temporal prediction is formed by the temporal interpolation of the current motion field and further refined by the motion field sent by the encoder.

6. A decoding device for carrying out a decoding method according to claim 5.

Abstract

The invention relates to an encoding method applied to an input video sequence corresponding to successive scenes subdivided into successive video object planes (VOPs) and generating, for coding all the video objects of said scenes, a coded bitstream constituted of encoded video data in which each data item is described by means of a bitstream syntax allowing to recognize and decode all the elements of the content of said bitstream. According to said method, said syntax comprises an additional syntactic element provided for describing independently the type of temporal prediction of the various channels, said additional syntactic element being placed at the image level in the coded bitstream and shared by all existing channels.

10

PCT/IB2004/001010

